

ANÁLISIS DE LA DIFICULTAD DE UN BANCO DE ÍTEMS DE VISUALIZACIÓN ESPACIAL

DIFFICULTY ANALYSIS IN A SPATIAL VISUALIZATION ITEM BANK

Gerardo Prieto

Universidad de Salamanca (España)

Angela D. Velasco

Universidade Estadual Paulista (Brasil)

Rosario Arias-Barahona

Universidad San Martín de Porres (Perú)

Mercedes Anido

Universidad Nacional de Rosario (Argentina)

Ana-María Núñez

Universidad San Martín de Porres (Perú)

Patricia Có

Universidad Nacional de Rosario (Argentina)

Resumen: Este trabajo describe el proceso de validación de un banco de ítems de Visualización basado en una tarea de Desarrollo de Superficies. El banco de ítems fue construido en el marco del diseño cognitivo. Los subtests fueron aplicados a muestras de estudiantes universitarios de ingeniería. Se estimó la dificultad de los ítems mediante el modelo de Rasch. Observamos que la dificultad está asociada a las condiciones de la tarea involucradas en la habilidad para transformar mentalmente las figuras y en el mantenimiento activo de la información espacial durante el proceso de transformación.

Palabras clave: *Visualización, Banco de ítems, Modelo de Rasch, Validez de los tests*

Abstract: This work is the description of the validation process of a Visualization item bank, whose items were generated from a Surfaces Development task. The item bank was developed from Cognitive Design framework. The tests were applied to samples of first-year Engineering students. The item difficulty was estimated using the Rasch model. We observed that the item difficulty is associated to a task conditions involved in the ability of mentally transforming the objects and of maintaining active the spatial information during the transforming process.

Key-words: *Visualization, Item bank, Rasch model, Test Validity*

INTRODUCCIÓN

La aplicación del diseño cognitivo a la construcción y validación de los ítems de los tests de aptitudes es una de las corrientes actuales más importantes en la medición de la inteligencia (Embretson, 1994; Irvine y Kyllonen, 2002). La utilización de principios cognitivos en el diseño requiere partir de unas hipótesis en las que se especifique la influencia de las características de los ítems en los procesos cognitivos, las

estrategias y las estructuras de conocimiento requeridas para la resolución de la tarea. En consecuencia, la generación de los ítems parte de un diseño en el que, por un lado, se seleccionan las características de los estímulos que hipotéticamente inducen los procesos relevantes para el constructo y, por otro, se controlan las características asociadas a los irrelevantes. El análisis revelará si las condiciones hipotetizadas están efectivamente asociadas a la dificultad empírica.

Este tipo de evidencia aporta indudables ventajas teóricas y tecnológicas. Desde la perspectiva teórica, constituye un medio de gran interés para analizar la validez de constructo de las medidas. Desde la perspectiva aplicada, es indudable que conocer las condiciones de la tarea vinculadas a la dificultad tiene importantes consecuencias para la construcción de tests. Entre las más destacables, cabe señalar la posibilidad de: (i) establecer reglas y procedimientos para definición del contenido y la generación de ítems (Bejar, 1993; Embretson, 2005), (ii) construir tests adaptativos informatizados a partir de criterios de contenido y (iii) equiparar el contenido de los tests en base a su complejidad cognitiva.

Desde el punto de vista metodológico, el análisis no debe realizarse sobre los resultados globales (puntuaciones en el test), sino sobre la respuesta a cada ítem. Por tanto, los modelos de la Teoría de Respuesta al Ítem (TRI) son los utilizados en este enfoque (Embretson, 2005).

Dentro de este contexto metodológico, el propósito general de este trabajo ha sido analizar la validez de un banco de ítems de Visualización que hemos desarrollado en los últimos años (...). En el próximo apartado se expondrán brevemente la metodología de validación basada en el Sistema de Diseño Cognitivo y la definición del constructo de Visualización que subyace a la construcción del banco de ítems.

Fundamentación teórica del problema

Para insertar el diseño de los tests en la metodología de validación de constructo, Embretson distinguió en 1983 dos aspectos en el concepto de validez: la representación del constructo y la amplitud nomotética. El primero se refiere a los procesos cognitivos, las estrategias y las estructuras de conocimiento requeridas para la resolución de la tarea correspondiente a los ítems. El segundo, más relacionado con los procedimientos tradicionales de validación, alude a las relaciones empíricas de las puntuaciones del test con otras medidas y a su utilidad para evaluar las diferencias individuales. El paradigma de investigación de la representación del constructo se enmarca en el uso de la metodología de la psicología cognitiva para construir modelos de procesamiento de la tarea y manipular las características de los ítems que demandan el empleo de los procesos relevantes para el constructo que se desea medir.

Esta metodología requiere modelos psicométricos que puedan incorporar las variables asociadas al diseño de los ítems. Como se ha indicado en la introducción, se emplean fundamentalmente variantes de los modelos TRI y en especial del modelo de Rasch (1960).

Este modelo, descrito brevemente en el apartado dedicado a la metodología, especifica que la probabilidad de una respuesta correcta a un ítem es una función de la diferencia entre el sujeto y el ítem en la variable medida. Se trata de un modelo muy atractivo para el análisis de la representación del constructo por varias razones:

1. Los parámetros de las personas y de los ítems se expresan en las mismas unidades y se localizan en el mismo continuo?

2. Se trata del único modelo TRI con la propiedad de *objetividad específica* (Rasch, 1977; Embretson y Reise, 2000). Esta propiedad implica que la diferencia de dificultad entre ítems no depende de las personas específicas que se utilicen para cuantificarla. En la Teoría Clásica de los Tests, por el contrario, la dificultad depende de la muestra de sujetos. En consecuencia, la obtención de conclusiones generalizables acerca de las fuentes de la dificultad de los ítems se logrará más fácilmente si se usa un modelo que permite obtener estimaciones libres de la muestra.

3. Inferir los componentes cognitivos, las estrategias y las estructuras de conocimiento asociados a la dificultad del ítem equivale a explicar el significado de la variable medida.

4. Facilita la calibración en la misma métrica de los subtests derivados del banco de ítems.

El estudio de las diferencias individuales en las aptitudes viso-espaciales se inició en el primer tercio del siglo veinte, debido a la utilidad de los tests espaciales para predecir el éxito en las profesiones técnicas. Los primeros investigadores en este campo (El Koussy, 1935; Kelly, 1928; Thurstone, 1938) efectuaron una clara distinción entre los componentes verbales y espaciales de la inteligencia humana, y verificaron que las aptitudes espaciales explicaban una cantidad muy importante de la varianza de todas las baterías multifactoriales de la inteligencia. En opinión de Hegarty y Waller (2005), esta conclusión sigue vigente en la actualidad. La Visualización (VZ) ha sido la aptitud espacial identificada de manera consistente en todos los estudios factoriales realizados en este campo. Guilford y Lacey (1947) la definieron inicialmente como la aptitud

para manipular, rotar, torcer o invertir figuras complejas. Sin duda, la Visualización es la aptitud espacial que ha recibido mayor atención debido a su importancia aplicada: existe un gran número de estudios que la identifican su eficacia predictora del rendimiento en ámbitos técnicos como la arquitectura, la ingeniería, el diseño gráfico, la cirugía, el pilotaje de aviones y el control aéreo (Ghiselli, 1973; Hsi, Linn y Bell, 1997; Murdoch, Bainbridge, Fisher y Webster, 1994; Peters, Chisholm y Laeng, 1995).

Mientras que los estudios factorialistas se focalizaron en la clasificación empírica y la descripción racional de las aptitudes viso-espaciales, los psicólogos cognitivos prefirieron analizar los procesos mentales demandados por dichas aptitudes. Desde este punto de vista, la ejecución de los tests espaciales depende de la puesta en ejercicio de procesos cognitivos básicos tales como la codificación de los estímulos visuales, la construcción de una imagen, su retención en la memoria de trabajo, la transformación mental de dicha imagen y la comparación del estímulo visual con la imagen almacenada en la memoria de trabajo. Desde los resultados de este marco de investigación cognitivo, se ha argumentado que las aptitudes espaciales identificadas por los factorialistas difieren en los mayores o menores requerimientos de estos procesos (Lohman, 2000). Por ejemplo, Pellegrino y Kail (1982) consideraron que las diferencias individuales en los tests más sencillos, como el PMA-E de Thurstone, pueden explicarse por las diferencias en la rapidez para ejecutar procesos elementales como la codificación perceptiva, la rotación mental y la comparación de figuras sencillas. Por el contrario, los ítems que saturan en el factor de Visualización demandan la habilidad para construir representaciones de figuras complejas y para mantener la calidad de la representación durante las transformaciones mentales sucesivas que se requieren en la solución de problemas. En consecuencia, Pellegrino y Kail (1982) afirmaron que si se ordenasen las aptitudes espaciales en los continuos de velocidad-potencia y de simplicidad-complejidad, la Visualización ocuparía la mayor jerarquía en potencia y complejidad.

Sin duda, la línea más importante de las investigaciones cognitivas sobre la aptitud de Visualización se refiere a su asociación con la Memoria de Trabajo. Existen evidencias de que

los sujetos con alto y bajo nivel de Visualización difieren en la calidad de las representaciones espaciales y en su habilidad para mantener dicha calidad durante el proceso de transformación mental de las figuras (Hegarty y Waller, 2005). Estos datos empíricos han sido interpretados en relación con las diferencias individuales en la memoria de trabajo, en el sentido de que los sujetos de alto nivel tienen más recursos para almacenar y procesar la información espacial que los de bajo nivel. Actualmente se concibe la memoria de trabajo como un sistema especializado en mantener la información activa y en controlar los recursos atencionales y los procesos requeridos durante la ejecución de una tarea compleja (Miyake y Shah, 1999). Asimismo, se cuenta con abundantes datos de la existencia de dos sistemas especializados para mantener activa la información verbal (phonological loop) y espacial (visual-spatial sketchpad) (Baddeley, 1986). Recientemente se ha propuesto que el sistema de Memoria de Trabajo Espacial, y en especial el componente ejecutivo, es el responsable de las diferencias individuales en la aptitud de Visualización (Hegarty, Shah y Miyake, 2000; Miyake, Rettinger, Friedman, Shah y Hegarty, 2001).

El banco de ítems de Visualización, cuya validez se pretende analizar, fue diseñado para evaluar las diferencias individuales en la ejecución de tareas que requieren tanto la representación y almacenamiento de figuras tridimensionales, como la ejecución de transformaciones mentales sucesivas de tipo espacial (rotación y plegamiento). Se manipularon las condiciones de la tarea para producir efectos en el tipo y el número de transformaciones y, en consecuencia, en su complejidad y dificultad. Se asume que las demandas a la Memoria de Trabajo Espacial se incrementan con el número de transformaciones requeridas por la tarea.

El objetivo específico de este trabajo ha sido analizar empíricamente las fuentes de la dificultad de los ítems con el fin de determinar la contribución de condiciones asociadas a la rotación mental (disparidad angular y eje de rotación) y al plegamiento mental de partes de la figura (lejanía del objetivo).

METODOLOGÍA

Instrumentos

Con el objetivo de minimizar las estrategias de resolución analítico-verbales, hemos optado

por usar en todos los ítems una figura regular. Cuando se emplean figuras irregulares, se facilita el etiquetado verbal de características distintivas de la figura (ángulos, tamaño, etc.) que permite el empleo de procesos que no son de carácter espacial. En consecuencia, la tarea consiste en un cubo en el que todas sus caras están identificadas con letras. A la derecha del cubo, aparece éste desplegado en el plano con una de sus caras identificada y otra marcada con una interrogación (?). Se pide al sujeto que identifique la letra de la cara marcada con la interrogación y su apariencia. El observador debe elegir la respuesta correcta entre las 9 opciones que se le facilitan (véase la Figura 1).

Las condiciones del dominio son las que siguen:

1. Rotación. Finalidad: inducir rotaciones mentales de la figura.

1.1. Disparidad angular. El cubo y la figura desplegada pueden aparecer con disparidades angulares de 90° y 180°.

1.2. Rotación en los siguientes ejes: X (vertical), Y (horizontal) y Z (profundidad).

2. Lejanía del objetivo (cara marcada con la interrogación) respecto de la cara identificada con una letra.

Para operacionalizar la condición 2, se designan las caras del cubo de la siguiente forma:

C0: Cara identificada con una letra en el cubo desplegado.

C1: Cara adyacente a C0.

C2: Cara adyacente a C1.

C3: Cara adyacente a C2.

C4: Cara adyacente a C3.

C5: Cara adyacente a C4.

Se incluyen ítems con la interrogación en las caras C2, C3, C4 y C5. Finalidad: inducir el número de plegamientos (torsiones) mentales

de partes de la figura. Se asume que el número de plegamientos se incrementará de acuerdo con el siguiente orden de las condiciones: C5>C4>C3>C2.

En la Figura 1 se muestra un ítem con las siguientes condiciones: Disparidad angular 90°, rotación los ejes Y y Z y lejanía del objetivo C4.

Se han calibrado 70 ítems distribuidos en seis tests: TVZ2002-A (20 ítems), TVZ2002-B (19 ítems), TVZ2002-C (20 ítems), TVZ2005-A (18 ítems), TVZ2005-B (18 ítems), TVZ2006A (20 ítems) y TVZ2006B (20 ítems). Los tests construidos en 2002 y 2005 eran de papel y lápiz. Los tests construidos en 2006 eran informatizados. Todos los tests tenían un conjunto de ítems comunes (anclaje) cuyo número oscilaba entre 4 y 8.

Participantes

Los tests se administraron a muestras de alumnos universitarios del primer curso de las licenciaturas de Ingeniería, Matemáticas y Medicina de universidades de Brasil (Universidade Estadual Paulista, Universidade de Taubaté, Faculdade de Engenharia Química de Lorena, Universidade de São Paulo, Instituto Tecnológico da Aeronáutica), Argentina (Universidad Nacional de Rosario) y Perú (Universidad San Martín de Porres). El tamaño y el país de procedencia se especifican a continuación.

TVZ2002-A: 238 alumnos de Brasil

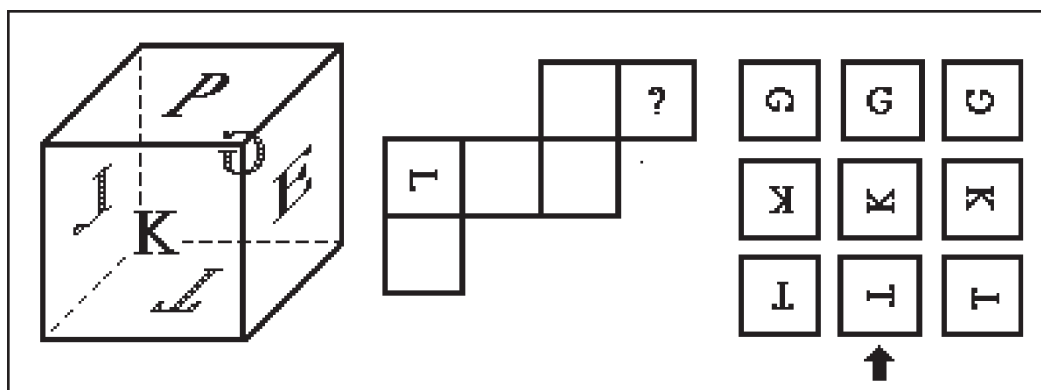
TVZ2002-B: 159 alumnos de Brasil

TVZ2002-C: 727 alumnos de Brasil y 591 de Perú

TVZ2005-A: 81 alumnos de Brasil y 116 de Argentina

TVZ2005-B: 56 alumnos de Brasil y 65 de Argentina

Figura 1. Modelo de ítem



TVZ2006-A: 149 alumnos de Argentina, 77 de Brasil y 171 de Perú

TVZ2006-B: 86 alumnos de Brasil, 149 de Argentina y 170 de Perú

Procedimiento

Los tests se administraron en las aulas habituales de cada grupo de alumnos. Una vez recibidas las instrucciones y los ejemplos de prueba, el tiempo de ejecución de los ítems fue de 20 minutos en los tests TVZ2002A y TVZ2002B y de 30 minutos en el resto de las pruebas.

Modelo de medida y de equiparación de puntuaciones

Como se ha indicado con anterioridad, se analizaron los datos mediante un modelo TRI muy conocido: el modelo logístico de un parámetro (Rasch, 1960). La ecuación básica (1) especifica que la probabilidad de resolver correctamente un ítem depende solamente de la diferencia en el atributo medido entre el nivel de la persona (θ_s) y el nivel del ítem (β_i). En este caso,

$$P_{is} = e^{(\theta_s - \beta_i)} / 1 + e^{(\theta_s - \beta_i)} \quad (1)$$

Donde e es la base de los logaritmos naturales (2,7183).

Los valores escalares de las personas y los ítems se sitúan en la misma escala. Estos valores pueden expresarse en distintas métricas (Embretson y Reise, 2000). La más utilizada es la escala *logit*, que es el logaritmo natural de $(P_{is} / 1 - P_{is})$. Aunque la escala *logit* puede adoptar valores entre más y menos infinito, la gran mayoría de los casos se sitúa en el rango ± 5 . La localización del punto 0 de la escala es arbitraria. En la tradición de Rasch, se suele situar dicho punto en la dificultad media de los ítems.

El modelo de Rasch tiene unas propiedades métricas óptimas: estadísticos suficientes, objetividad específica, unidimensionalidad, métrica intervalar, parámetros independientes de la muestra y estadísticos para analizar el ajuste al modelo de ítems y personas (Santisteban y Alvarado, 2001). Aunque la funcionalidad de estas ventajas depende del ajuste de los datos a los supuestos del modelo, es conveniente notar que basta con un ajuste razonable para justificar su uso: los estudios acerca de la robustez del modelo de Rasch ante las violaciones de sus supuestos muestran que las ganan-

cias producidas por el uso de otros modelos más complejos son escasas (Muñiz, 1997).

Como indicador del ajuste de ítems y personas se han empleado los estadísticos Infit y Outfit, que son medias de los residuos estandarizados al cuadrado (diferencias entre el valor observado y el predicho por el modelo). La fórmula de un residuo es:

$$y_{is} = (x_{is} - P_{is}) \quad (2)$$

Donde x_{is} es la respuesta observada y P_{is} la probabilidad de una respuesta correcta de la persona s al ítem i . Se suelen estandarizar los residuos dividiéndolos por su desviación típica:

$$z_{is} = (x_{is} - P_{is}) / ((P_{is}(1 - P_{is}))^{1/2}) \quad (3)$$

Infit es la media de los residuos cuadráticos ponderados con su varianza (W_{is}).

$$\text{Infit} = \sum z_{is}^2 W_{is} / \sum W_{is} \quad (4)$$

Outfit es la media de los residuos cuadráticos no ponderados.

$$\text{Outfit} = \sum z_{is}^2 / N \quad (5)$$

El valor esperado de estos estadísticos es 1. Por convención se considera que los valores superiores a 1,5 indican un desajuste moderadamente alto, y que los valores superiores a 2,0 revelan muy alto desajuste (Wright y Linacre, 1998).

Outfit es un indicador muy sensible a los outliers (basta una respuesta muy inesperada para que adopte un valor muy elevado). Infit es más robusto: los valores altos se deben a patrones de respuesta claramente aberrantes.

Como medida de la fiabilidad de las puntuaciones se ha empleado un análogo del coeficiente de fiabilidad clásico: el índice de fiabilidad (6). Este estadístico se calcula por separado para las personas y para los ítems.

$$\text{Fiabilidad} = \text{varianza observada} - \text{varianza de error} / \text{varianza observada} \quad (6)$$

La varianza observada es la varianza de los parámetros de los sujetos o los ítems. La varianza de error es el cuadrado del promedio de los errores estándar de los sujetos o los ítems.

Se empleó un procedimiento de equiparación de la métrica del banco utilizando un diseño de anclaje con los ítems comunes con el test TVZ2002-C (test de referencia), debido a que el gran tamaño de la muestra utilizada en la calibración de este test permitía la mejor estimación de los parámetros de los ítems de anclaje. El procedimiento de escalamiento común de los

ítems de los cinco tests consiste en calibrar los datos de cada test por separado, analizar la invarianza de los parámetros de dificultad de los ítems comunes entre dos tests (el test de referencia y el test anclado) mediante la correlación de Pearson y, en el caso de que sean invariantes, volver a estimar los parámetros de los ítems del test anclado fijando la dificultad de los ítems de anclaje en el test de referencia. Una descripción más exhaustiva del procedimiento puede consultarse en ...

El análisis del ajuste de los datos al modelo de Rasch, la estimación de los parámetros de los ítems y el establecimiento de una métrica común para los ítems de los cinco tests (equiparación) se llevó a cabo mediante el programa Winsteps (Wright y Linacre, 1998).

RESULTADOS

En la Tabla 1 aparecen los principales valores psicométricos de los ítems de los siete tests, una vez establecida una métrica común para todo el banco. Se puede considerar que el procedimiento de equiparación de los tests es adecuado, dada la invarianza entre los ítems de anclaje (las correlaciones entre los parámetros de dificultad de los ítems de anclaje oscilaron

entre 0,87 y 1,00). Asimismo, son muy elevadas las correlaciones entre las calibraciones por separado de los ítems y sus parámetros en el banco (entre ,88 y 1,00).

Puede observarse que hay un buen ajuste de los ítems al modelo: sólo 7 ítems de los 70 que componen el banco presentan valores de infit y/o outfit mayores de 1,5 (ninguno de ellos eran ítems de anclaje). La fiabilidad de los parámetros de dificultad de los ítems es elevada (los índices de fiabilidad de los ítems en los cinco tests oscilan entre ,81 y ,99). Efectuado una transformación lineal de la escala para situar el punto cero en la dificultad media de los ítems del banco (-,25), se observa que el rango de dificultad de sus parámetros oscila entre -2,17 y 1,68 y que la distribución es asimétrica, siendo más abundantes los ítems difíciles (Figura 2).

Como ya se ha indicado, el objetivo de este trabajo ha sido indagar los efectos en la dificultad de las condiciones de los ítems. Para esclarecer este punto, se llevó a cabo un análisis de regresión múltiple en el que las estimaciones de la dificultad de los ítems fue la variable dependiente (β^*) y los predictores, la disparidad angular (90° y 180°), el número de ejes de rotación y la lejanía del objetivo (C2, C3, C4 y C5).

Tabla 1. Valores psicométricos de los ítems de los tests

Test	TVZ02A	TVZ02B	TVZ02C	TVZ05A	TVZ05B	TVZ06A	TVZ06B
Ítems	20	19	20	18	18	20	20
N	238	159	1313	192	192	397	405
Media β	-,86	-,06	,00	-,26	-,04	-,40	,25
DT β	1,17	,69	,65	,68	,71	,70	,58
Máxima β	1,25	1,18	1,01	,67	,84	,67	1,18
Mínima β	-2,42	-1,27	-1,27	-2,18	-2,18	-2,18	-1,52
Media SE (β)	,20	,28	,07	,21	,26	,13	,15
DT SE (β)	,02	,07	,01	,01	,01	,01	,01
Fiabilidad	,97	,81	,99	,90	,87	,97	,93
Media Infit (I)	,96	1,00	,99	1,00	,99	1,00	1,02
DT Infit (I)	,15	,21	,08	,11	,14	,13	,14
Infit máximo	1,24	1,70	1,15	1,21	1,31	1,24	1,39
Infit mínimo	,75	,74	,83	,77	,69	,75	,82
Ítems Infit > 1,5	0	1	0	0	0	0	0
Media Outfit (I)	1,06	1,07	1,03	1,02	1,03	1,02	1,02
DT Outfit (I)	,43	,50	,14	,20	,26	,23	,23
Outfit máximo	2,02	2,79	1,27	1,47	1,64	1,44	1,63
Outfit mínimo	,38	,52	,73	,65	,58	,65	,72
Ítems I Outfit > 1,5	3	2	0	0	1	0	1
Ítems anclaje	4	6	--	8	5	8	7
R ítems anclaje	,97	,97	--	,89	1,00	,91	,87
R β β^*	,98	,95	1,00	,97	,96	,94	,88

N=número de casos

β =parámetros de los ítems (calibración por separado)

β^* = parámetros de los ítems (banco)

La correlación entre la dificultad empírica (β) y la dificultad predicha (β^*) fue de 0,72. El coeficiente de determinación múltiple ($R^2=0,51$) indica la proporción de la varianza de la dificultad empírica explicable por la combinación lineal de los predictores. Se puede concluir que la cantidad explicada es elevada.

En la Tabla 2, aparecen diversos estadísticos que permiten inferir la contribución diferencial de los predictores a la varianza de la dificultad de los ítems: los coeficientes estandarizados de regresión múltiple (Tabachnick y Fidell, 2007).

Puede observarse que sólo son significativos los efectos de las condiciones denominadas *lejanía del objetivo* y *rotación*, siendo la primera el predictor más eficiente. Por el contrario, el número de ejes de rotación no parece tener una contribución relevante sobre la dificultad.

Para analizar las interacciones entre estas dos condiciones que son más relevantes, se realizó un análisis de varianza factorial (Tabla 3).

Los efectos de la Rotación y de la Lejanía y su interacción son estadísticamente significativos. Un gráfico de las líneas de interacción facilita la interpretación de estos efectos (Figura 3).

Figura 2. Distribución de la dificultad de los ítems

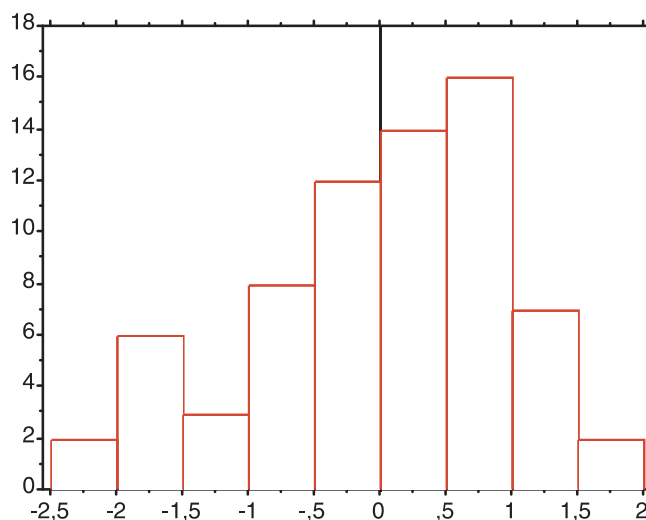


Tabla 2. Efecto en la dificultad de los ítems de las condiciones de la tarea

Factor	β	t	p
Rotación	,27	2,15	,0349
Lejanía	,57	6,44	,0001
Ejes	,10	,82	,4154

Beta = Coeficiente estandarizado de regresión parcial

t= t de Student

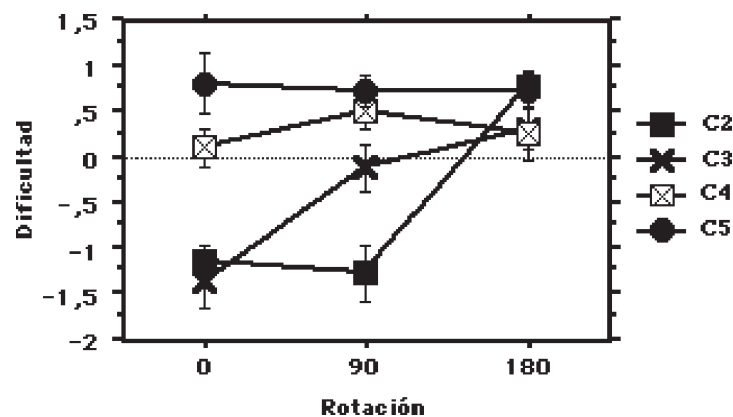
p= significación

Coeficiente de determinación múltiple = 0,49 ($R=0,70$)

Tabla 3. ANOVA. Variable dependiente: Dificultad de los ítems. Factores: Rotación y Lejanía

Factor	GL	Suma de cuadrados	Media de cuadrados	F	P
Rotación	2	8,23	4,12	10,17	,0002
Lejanía	3	14,41	4,8	11,87	,0001
Interacción	6	7,97	1,33	3,28	,0077
Residual	57	23,06	,40		

Figura 3. Interacción entre la Rotación y la Lejanía



Obsérvese que los mayores niveles de dificultad de los ítems se alcanzan con los mayores valores de Lejanía (C4 y C5), independientemente de la rotación de las figuras. La condición de Rotación incrementa la dificultad en los niveles bajos de Lejanía. Estos datos esclarecen aún más la importancia de esta condición.

DISCUSIÓN

En este trabajo se ha analizado la influencia en la variable medida, cuantificada mediante los parámetros de dificultad de los ítems, de las características de la tarea de un banco de ítems de Visualización Espacial. Para estimar estos parámetros se empleó el modelo de Rasch. Los resultados indican que la condición denominada *Lejanía del Objetivo* es la que explica más consistentemente las variaciones de nivel en la variable. Sin embargo, la condición de *Rotación* tiene un efecto más leve, incrementando moderadamente la dificultad en las figuras con bajos valores de Lejanía. Estos datos son convergentes con los análisis de otras tareas de desplegamiento como la del test DAT-SR (Bennett, Seashore y Wesman, 1951) y la del test SLAT (Embretson, 1994), en el sentido de que las condiciones que inducen transformaciones espaciales como plegamientos o desplegamientos y torsiones de las partes de la figura son las más definitorias de los tests de Visualización.

Estos hallazgos son útiles para esclarecer el significado de la tarea (la representación del constructo). En los últimos años se ha puesto el énfasis en la importancia de la memoria de trabajo espacial. Desde este punto de vista, el mantenimiento activo durante las transformaciones de las características de la figura y el control de las transformaciones jugarían un papel relevante

en las tareas de Visualización (Logie, 1995; Miyake y Shah, 1999). Parece razonable suponer que la manipulación de la lejanía del objetivo y de la rotación afectaría a las demandas a la memoria de trabajo visual. No obstante, es necesario estudiar la amplitud nomotética de esta tarea analizando las correlaciones con marcadores externos de Memoria de Trabajo Espacial.

Desde la perspectiva aplicada de la construcción de instrumentos de evaluación, los datos aportados por este trabajo pueden ser útiles para definir el contenido de futuros tests, establecer reglas tendentes a la generación automática de ítems, construir tests adaptativos informatizados a partir de criterios de contenido y equiparar tests en base a su complejidad cognitiva.

REFERENCIAS

- Baddeley, A. D. (1986). *Working Memory*. New York: Oxford University Press.
- Bejar, I.I. (1993). A Generative Approach to Psychological and Educational Measurement. En N. Frederiksen, R.J. Mislevy e I.I. Bejar (Eds.), *Test Theory for a new generation of tests*. (pp. 323-358). Hillsdale, N.J.: LEA.
- Bennett, G. K., Seashore, H. G. y Wesman, A. G. (1951). *Counseling from profiles: A casebook for the Differential Aptitude Tests*. New York: Psychological Corporation.
- El Koussy, A. A. H. (1935). An investigation into the factors involving the visual perception of space. *British Journal of Psychology*, 20, 92.
- Embretson, S.E. (1983). Construct validity: Construct representation versus nomothetic span. *Psychological Bulletin*, 93, 179-197.
- Embretson, S. E. (1994). Applications of Cognitive Design Systems to Test Development. En C.R. Reynolds (Ed.), *Cognitive Assessment: A Multidisciplinary Perspective*. (pp. 107-135).

- Embretson, S. E. (2005). Measuring Human Intelligence with Artificial Intelligence. En R. J. Sternberg y J. E. Pretz (Eds.) *Cognition & Intelligence: Identifying the Mechanisms of the Mind*. (pp. 251-267). New York: Cambridge University Press.
- Embretson, S.E. y Reise, S.P. (2000) *Item Response Theory for Psychologists*. Mahwah, NJ: LEA.
- Guilford, J. P. y Lacey, J. I. (1947). *Printed Classification Tests*, A. A. F. Aviation Psychological Progress Research Report, No. 5. Washington, DC: U. S. Government Printing Office.
- Ghiselli, E. E. (1973). The validity of aptitude tests in personnel selection. *Personnel Psychology*, 26, 461-477.
- Hegarty, M., Shah, P. Y Miyake, A. (2000). Constraints on using the dual-task methodology to specify the degree of central executive involvement in cognitive tasks. *Memory and Cognition*, 28, 411-430.
- Hegarty, M. y Waller, D.A. (2005). Individual Differences in Spatial Abilities. En P. Shah y A. Miyake (Eds.). *The Cambridge Handbook of Visuospatial Thinking*. (pp. 121-169). New York: Cambridge University Press.
- Hsi, S., Linn, M.C. e Bell, J.E. (1997). The role of Spatial Reasoning in Engineering and Design of Spatial Instruction. *Journal of Engineering Education*, 84, 151-158.
- Irvine, S.H. y Kyllonen, P. (2002). *Item generation for test development*. Mahwah, NJ: LEA.
- Kelly, T. L. (1928). *Crossroads in the mind of man*. Stanford: Stanford University Press.
- Logie, R.H. (1995). *Visuo-Spatial Working Memory*. Hove: LEA.
- Lohman, D.F. (2000). Complex Information Processing and Intelligence. En R.J. Sternberg (Org.) *Handbook of Intelligence*. (pp. 285-340). Cambridge, UK: Cambridge University Press.
- Miyake, A. y Shah, P. (Eds.) (1999). *Models of working memory*. Cambridge: Cambridge University Press.
- Miyake, A., Rettinger, D. A., Friedman, N. P., Shah, P. Y Hegarty, M. (2001). Visuospatial working memory, executive functioning and spatial abilities. How are they related? *Journal of Experimental Psychology: General*, 130, 621-640.
- Muñiz, J. (1997). *Introducción a la Teoría de Respuesta a los Ítems*. Madrid: Pirámide.
- Murdoch, J. R., Bainbridge, L. C., Fisher, S. G. y Webster, M. H. C. (1994). Can a simple test of visual motor skill predict the performance of microsurgeons? *Journal of the Royal College of Surgeons, Edinburg*, 39, 150-152.
- Pellegrino, J.W. & Kail, R. (1982) *Process Analyses of Spatial Aptitude*. En R.J. Sternberg (Ed.), *Advances in the psychology of human intelligence*, 1. (pp. 311-365). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Peters, M., Chisholm, P. y Laeng, B. (1995). Spatial Ability, Student Gender, and Academic Performance. *Journal of Engineering Education*, 84, 69-73.
- Prieto, G. y Delgado, A.R. (2002). Diseño cognitivo de un banco de ítems de visualización espacial. *Metodología de las Ciencias del Comportamiento*, volumen especial, 452-455.
- Prieto, G. y Velasco, A. D. (2002a). Construção de um teste de visualização a partir da psicologia cognitiva. *Avaliação Psicologica*, 1, 39-47.
- Prieto, G. e Velasco, A.D. (2002b). Predicting Academic Success of Engineering Students in Technical Drawing from Visualization Test Scores. *Journal for Geometry and Graphics*, 6, 99-109.
- Prieto, G. y Velasco, A.D. (2003). Uso del modelo de Rasch para poner en la misma escala las puntuaciones de distintos tests. *Actualidades en Psicología*, 19, 7-24.
- Prieto, G. y Velasco, A. D. (2004). Training Visualization Ability by Technical Drawing. *Journal for Geometry and Graphics*, 8, 107-115.
- Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen: The Danish Institute for Educational Research.
- Santisteban, C. y Alvarado, J.M. (2001). *Modelos psicométricos*. Madrid: UNED.
- Tabachnick, B.G. y Fidell, L.S. (2007). *Using multivariate statistics*. Boston: Pearson.
- Thurstone, L. L. (1938). *Primary Mental Abilities*. Chicago: University of Chicago Press.
- Ting, S.R. (2001). Predicting Academic Success of First-Year Engineering Students from Standardized Test Scores and Psychosocial Variables. *International Journal of Engineering Education*, 17, 75-80.
- Wright, B.D. y Linacre, J.M. (1998). *WINSTEPS : A Rasch computer program*. Chicago: MESA Press.